## Wykład 4: Podstawowe własności estymatorów

W wykładzie tym przedstawimy kryteria pozwalające ocenić jakość danego estymatora, bądź wybrać z pewnego zbioru ten, który w pewnym sensie jest najlepszy.

### Estymatory nieobciążone

Zacznijmy od wprowadzenia definicji.

**Definicja.** Mówimy, że $\hat{\theta} = t(X_1, X_2, \ldots, X_n)$ jest *nieobciążonym estymatorem* parametru $\theta$ jeśli

$$E_\theta(\hat{\theta}) = \theta \quad \text{dla każdego } \theta \in \Theta.$$

W pozostałych przypadkach $\hat{\theta}$ nazywamy *estymatorem obciążonym.* Funkcję $B(\theta) = E_\theta(\hat{\theta}) - \theta$, gdzie $\theta \in \Theta$, nazywamy *obciążeniem estymatora* $\hat{\theta}$.

Z powyższych definicji natychmiasyt wynika, że $\hat{\theta}$ jest estymatorem nieobciążonym wtedy i tylko wtedy, gdy jego obciążenie jest funkcją tożsamościowo równą zero.

**Przykład 4.1.** Niech $X_1, X_2, \ldots, X_n$ będzie prostą próbą losową z populacji $X$ o rozkładzie z wartością oczekiwaną $EX = \mu$. Wówczas $\hat{\mu} = \bar{X}$ jest nieobciążonym estymatorem parametru $\mu$. Rzeczywiście dla dowolnego $\mu$ otrzymujemy

$$E_\mu(\hat{\mu}) = E_\mu(\bar{X}) = E_\mu\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n}\sum_{i=1}^n E_\mu(X_i) = \frac{1}{n}\sum_{i=1}^n \mu = \frac{1}{n}n\mu = \mu.$$

Dla $X$ o rozkładzie dwupunktowym z prawdopodobieństwem sukcesu $p$ mamy $EX = p$. Zatem jako szczególny przypadek powyższego wyniku otrzymujemy, że jeśli $X_1, X_2, \ldots, X_n$ jest prostą próbą losową z populacji $X$ o rozkładzie dwupunktowym z prawdopodobieństwem sukcesu $p$, to $\hat{p} = \bar{X}$ jest nieobciążonym estymatorem parametru $p$.

**Przykład 4.2.** Niech $X_1, X_2, \ldots, X_n$ będzie prostą próbą losową z populacji $X$ o rozkładzie z wartością oczekiwaną $EX = \mu$ i wariancją $Var(X) = \sigma^2 > 0$. Załóżmy, że nie znamy ani $\mu$ ani $\sigma^2$. Wówczas

$$\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$$

jest estymatorem obciążonym parametru $\sigma^2$, bo

$$E_{(\mu,\sigma^2)}(\hat{\sigma^2}) = \frac{n-1}{n}\sigma^2 \neq \sigma^2,$$

natomiast

$$\hat{\sigma^2_{no}} = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$$

jest estymatorem nieobciążonym $\sigma^2$.

Aby to pokazać zauważmy najpierw, że

$$\begin{aligned}\sum_{i=1}^n (X_i - \bar{X})^2 &= \sum_{i=1}^n (X_i - \mu - (\bar{X} - \mu))^2 = \sum_{i=1}^n (X_i - \mu)^2 - 2(\bar{X} - \mu)\sum_{i=1}^n (X_i - \mu) + n(\bar{X} - \mu)^2 = \\ &= \sum_{i=1}^n (X_i - \mu)^2 - 2n(\bar{X} - \mu)^2 + n(\bar{X} - \mu)^2 = \sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X} - \mu)^2,\end{aligned}$$

bo $\sum_{i=1}^n (X_i - \mu) = n \cdot \frac{1}{n}\sum_{i=1}^n X_i - n\mu = n(\bar{X} - \mu)$. Stąd

$$\begin{aligned} E_{(\mu,\sigma^2)}(\hat{\sigma^2}) &= E_{(\mu,\sigma^2)}\left(\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2\right) = \frac{1}{n}E_{(\mu,\sigma^2)}\left(\sum_{i=1}^n (X_i-\mu)^2 - n(\bar{X}-\mu)^2\right) \\ &= \frac{1}{n}\left[E_{(\mu,\sigma^2)}\left(\sum_{i=1}^n (X_i-\mu)^2\right) - nE_{(\mu,\sigma^2)}(\bar{X}-\mu)^2\right] = \frac{1}{n}(n\sigma^2 - \sigma^2) = \sigma^2\frac{n-1}{n}, \end{aligned}$$

ponieważ

$$E_{(\mu,\sigma^2)}(X_i - \mu)^2 = E_{(\mu,\sigma^2)}\left(X_i - E_{(\mu,\sigma^2)}(X_i)\right)^2 = Var_{(\mu,\sigma^2)}(X_i) = \sigma^2$$

oraz, zgodnie z tym, co pokazaliśmy w przykładzie 4.1, $E_\mu(\bar{X}) = \mu$, co pociąga za sobą

$$\begin{aligned} E_{(\mu,\sigma^2)}\left(\bar{X}-\mu\right)^2 &= E_{(\mu,\sigma^2)}\left(\bar{X} - E_{(\mu,\sigma^2)}(\bar{X})\right)^2 = Var_{(\mu,\sigma^2)}(\bar{X}) = Var_{(\mu,\sigma^2)}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \\ &= \frac{1}{n^2}\sum_{i=1}^n Var_{(\mu,\sigma^2)}(X_i) = \frac{1}{n^2}\cdot n\sigma^2 = \frac{1}{n}\sigma^2. \end{aligned}$$

Wprowadzając drobną modyfikację w mianowniku estymatora $\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$, przerobimy go na estymator nieobciążony. Rzeczywiście

$$\begin{aligned} E_{(\mu,\sigma^2)}\left(\frac{1}{n-1}\sum_{i=1}^n (X_i-\bar{X})^2\right) &= E_{(\mu,\sigma^2)}\left(\frac{n}{n-1}\frac{1}{n}\sum_{i=1}^n (X_i-\bar{X})^2\right) = \frac{n}{n-1}E_{(\mu,\sigma^2)}\left(\hat{\sigma^2}\right) \\ &= \frac{n}{n-1}\frac{n-1}{n}\sigma^2 = \sigma^2, \end{aligned}$$

co pokazuje, że $\hat{\sigma^2_{no}} = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$ jest estymatorem nieobciążonym $\sigma^2$.

Choć, w przypadku próby losowej z rozkładu o istniejących ale nieznanych wartości oczekiwanej $\mu$ i wariancji $\sigma^2$, estymator $\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$ jest obciążonym estymatorem $\sigma^2$, to wraz ze wzrostem liczności próby jego obciążenie maleje do zera:

$$\lim_{n\to\infty} E_{(\mu,\sigma^2)}(\hat{\sigma^2}) = \lim_{n\to\infty}\frac{n-1}{n}\sigma^2 = \sigma^2,$$

a stąd

$$\lim_{n\to\infty} B(\hat{\sigma^2}) = \lim_{n\to\infty}(\sigma^2 - E_{(\mu,\sigma^2)}(\hat{\sigma^2})) = 0.$$

Estymator o takiej własności (a precyzyjniej - ciąg estymatorów $\hat{\theta}_n$, gdzie $n$ to liczność próby losowej) nazywamy asymptotycznie nieobciążonym.

**Definicja.** Mówimy, że ciąg estymatorów $\hat{\theta}_n = t_n(X_1, X_2, \ldots, X_n)$ parametru $\theta$ jest *asymptotycznie nieobciążony* jeśli

$$\lim_{n\to\infty} E_\theta(\hat{\theta}_n) = \theta \quad \text{dla każdego } \theta \in \Theta.$$

### Estymatory nieobciążone o minimalnej wariancji

Nieobciążoność danego estymatora nie gwarantuje nam, że jest on dobry - przyjmowane przez niego wartości mogą być bardzo rozproszone wokół $\theta$. Bardziej porządaną własnością od nieobciążoności jest by rozproszenie estymatora wokół szacowanej $\theta$ było możliwie najmniejsze. Rozproszenie to możemy mierzyć jako wartość oczekiwana z kwadratu różnicy między estymatorem i $\theta$.

**Definicja.** Funkcję

$$MSE_{\hat{\theta}}(\theta) = E_\theta \left(\hat{\theta} - \theta\right)^2, \quad \text{gdzie } \theta \in \Theta,$$

nazywamy *błędem średniokwadratowym estymatora* $\hat{\theta}$ parametru $\theta$.

Najlepszym estymatorem byłby estymator minimalizujący błąd średniokwadratowy dla wszystkich $\theta \in \Theta$. Niestety, poza wyjątkowo rzadkimi przypadkami, taki estymator nie istnieje. Jeśli jednak ograniczymy zbiór rozważanych estymatorów do nieobciążonych, to sytuacja ulegnie zmianie.

Zauważmy, że

$$\begin{aligned} MSE_{\hat{\theta}}(\theta) &= E_\theta \left(\hat{\theta} - \theta\right)^2 = E_\theta \left(\hat{\theta} - E_\theta\hat{\theta} - (\theta - E_\theta\hat{\theta})\right)^2 \\ &= E_\theta \left(\hat{\theta} - E_\theta\hat{\theta}\right)^2 - 2(\theta - E_\theta\hat{\theta})E_\theta\left(\hat{\theta} - E_\theta\hat{\theta}\right) + (\theta - E_\theta\hat{\theta})^2 = \\ &= Var_\theta(\hat{\theta}) + (B(\theta))^2, \end{aligned} \tag{1}$$

bo $E_\theta\left(\hat{\theta} - E_\theta\hat{\theta}\right) = E_\theta\hat{\theta} - E_\theta\left(E_\theta\hat{\theta}\right) = E_\theta\hat{\theta} - E_\theta\hat{\theta} = 0$. Oznacza to, że błąd średniokwadratowy estymatora to suma jego wariancji i kwadratu obciążenia. Błąd ten zatem będzie mały jedynie, gdy zarówno wariancja estymatora jak i jego obciążenie będą małe. Natomiast dla nieobciążonych estymatorów $\hat{\theta}$ parametru $\theta$ wzór (1) redukuje się do

$$MSE_{\hat{\theta}}(\theta) = Var_\theta(\hat{\theta}),$$

i minimalizowanie błędu średniokwadratowego jest równoważne minimalizowaniu wariancji.

**Definicja.** Estymator $T_0 = t(X_1, \ldots, X_n)$ nazywamy *estymatorem nieobciążonym o minimalnej wariancji* parametru $\theta$, jeśli

1. $T_0$ jest nieobciążony, tzn. dla każdego $\theta \in \Theta$ mamy $E_\theta(T_0) = \theta$,
2. $Var_\theta(T_0) \leq Var_\theta(T)$ dla każdego $\theta \in \Theta$ i dla każdego estymatora nieobciążonego $T$ parametru $\theta$.

W literaturze anglosaskiej estymator nieobciążony o minimalnej wariancji w skrócie nazywa się UMVUE od *uniformly minimum-variance unbiased estimator*.

## Estymatory zgodne

Niech $X_1, X_2, \ldots$ będzie ciągiem niezależnych zmiennych losowych o tym samym rozkładzie zależnym od parametru $\theta \in \Theta$. Dla dowolnego naturalnego $n$ tworzymy próbę losową $X_1, X_2, \ldots, X_n$ i na jej podstawie budujemy estymator $\hat{\theta}_n = t_n(X_1, X_2, \ldots, X_n)$ parametru $\theta$.

**Definicja.** Mówimy, że ciąg estymatorów $\hat{\theta}_n = t_n(X_1, X_2, \ldots, X_n)$, $n \geq 1$, parametru $\theta$ jest

- *zgodny w sensie zbieżności średniokwadratowej* jeśli błąd średniokwadratowy $\hat{\theta}_n$ zbiega do zera wraz ze wzrostem liczności próby do nieskończoności:

$$\lim_{n\to\infty} E\left((\hat{\theta}_n - \theta)^2\right) = 0 \qquad \text{dla wszystkich } \theta \in \Theta;$$

- *mocno zgodny* jeśli z prawdopodobieństwem 1 realizacje $\hat{\theta}_n$ dążą do $\theta$, gdy liczność próby wzrasta do nieskończoności

$$P\left(\lim_{n\to\infty} \hat{\theta}_n = \theta\right) = 1 \qquad \text{dla wszystkich } \theta \in \Theta;$$

- *(słabo) zgodny* jeśli dla dostatecznie dużych liczności próby estymator $\hat{\theta}_n$ z dużym prawdopodobieństwem przyjmuje wartości bliskie $\theta$:

$$\lim_{n\to\infty} P\left(|\hat{\theta}_n - \theta| < \varepsilon\right) = 1 \quad \text{dla każdego } \varepsilon > 0 \text{ i dla wszystkich } \theta \in \Theta.$$

W terminach rodzajów zbieżności ciągów zmiennych losowych

- zgodność w sensie zbieżności średniokwadratowej oznacza, że ciąg $\hat{\theta}_n$ zbiega w sensie zbieżności średniokwadratowej do $\theta$: $\hat{\theta}_n \xrightarrow{\mathrm{L}^2} \theta$;
- mocna zgodność oznacza, że ciąg $\hat{\theta}_n$ zbiega z prawdopodobieństwem 1 (prawie na pewno) do $\theta$: $\hat{\theta}_n \xrightarrow{1} \theta$;
- słaba zgodność oznacza, że ciąg $\hat{\theta}_n$ zbiega według prawdopodobieństwa do $\theta$: $\hat{\theta}_n \xrightarrow{P} \theta$.

Z terorii rachunku prawdopodobieństwa wynika, że każdy estymator zgodny w sensie zbieżności średniokwadratowej jak i każdy estymator mocno zgodny jest zgodny.

Podane niżej twierdzenie przyda nam się do uzasadnienia, że średnia próbkowa jest mocno zgodnym estymatorem średniej populacyjnej.

**Twierdzenie 4.1** (Mocne prawo wielkich liczb Kołmogorowa)**.** Jeśli $X_1, X_2, \ldots$ jest ciągiem niezależnych zmiennych losowych o tym samym rozkładzie ze skończoną wartością oczekiwaną $\mu$, to

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{1} \mu \quad \text{tzn.} \quad P\left(\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = \mu\right) = 1.$$

**Przykład 4.3.** Niech $X_1, X_2, \ldots$ będzie ciągiem niezależnych zmiennych losowych o tym samym rozkładzie z nieznaną wartością oczekiwaną $\mu$. Z mocnego prawa wielkich liczb Kołmogorowa natychmiast wynika, że $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ jest mocno zgodnym estymatorem parametru $\mu$. Stąd $\bar{X}_n$ jest też słabo zgodnym estymatorem $\mu$.

Jeśli dodatkowo założymy, że $X_1, X_2, \ldots$ mają wariancję $\sigma^2$, to otrzymamy, że $\bar{X}_n$ jest także zgodnym w sensie zbieżności średniokwadratowej estymatorem $\mu$. Rzeczywiście, skoro $\mu$ jest nieobciążonym estymatorem $\mu$, to otrzymujemy

$$E\left((\bar{X}_n - \mu)^2\right) = E\left((\bar{X}_n - E(\bar{X}_n))^2\right) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n} \xrightarrow{n\to\infty} 0.$$

Następny przykład podamy bez dowodu.

**Przykład 4.4.** Niech $X_1, X_2, \ldots$ będzie ciągiem niezależnych zmiennych losowych o tym samym rozkładzie z nieznanymi wartością oczekiwaną $\mu$ i wariancją $\sigma^2$. Wtedy

$$\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X})^2 \quad \text{i} \quad \hat{\sigma^2_{no}} = \frac{1}{n-1}\sum_{i=1}^{n} (X_i - \bar{X})^2$$

są mocno zgodnymi estymatorami parametru $\sigma^2$. Zatem są to także estymatory słabo zgodne. Ponadto, przy dodatkowym założeniu, że $X_1, X_2, \ldots$ mają skończony czwarty moment centralny, zarówno $\hat{\sigma^2}$ jak i $\hat{\sigma^2_{no}}$ są zgodnymi w sensie zbieżności średniokwadratowej estymatorami parametru $\sigma^2$.